ENERGY SAVING INFORMATION CASCADES IN ONLINE SOCIAL NETWORKS: AN AGENT-BASED SIMULATION STUDY

Qi Wang
John E. Taylor

Dept. of Civil & Env. Eng.
Virginia Tech
Blacksburg, VA 24061, USA

ABSTRACT

Information shared through online social networking platforms is spread from user to user. Although some researchers have argued that this phenomenon can unfold similarly to an epidemic, others have found that information disseminates within a narrow range, propagating only a few levels in a communication network. In an effort to resolve these conflicting findings, we developed an information cascade model to conduct a variance-based global sensitivity analysis (GSA) to determine the influence of two network attributes on the diffusion of energy saving information. The simulation results of the base model showed that energy saving information failed to generate deep cascades. Also, the results from the GSA demonstrated that network density and the number of an initiator's connections had limited influence on information cascades. These findings suggest that massive network structures and a large number of potential recipients do not engender deep cascades of energy saving information in online social networks.

1   INTRODUCTION

The U.S. is facing the grand challenge of energy independence. The *Energy Information Administration* (2011) reports that the energy consumption has tripled between 1949 and 2011. The increased demand has led to heavy dependence on imports. For example, 58% of petroleum consumed in the U.S. in 2007 came from outside its borders (National Science Foundation 2009). In order to achieve energy independence, reducing energy consumption is as important as finding new energy resources.

Reducing energy consumption requires both advances in technology and changes in people's behavior. Turner and Frankel (2008) studied 100 LEED (Leadership in Energy and Environmental Design) certified buildings, comparing their actual energy performance to projected performance during the design phase. They found that not only did almost 33% of the buildings fall below their design standard, but over 13% performed below the code baseline, largely due to their occupants' behaviors. Recent work has concluded that occupant behavior can significantly impact energy use in buildings (Azar and Menassa 2012; Masoso and Grobler 2010). Clearly, if the goal of energy independence is to be achieved, building occupants will need to change their energy consumption decisions and behaviors.

To accomplish a change of this magnitude, it is necessary to provide users with more energy saving related information. Wilson and Dowlatabadi (2007) observed that energy conservation behaviors and technologies often remain underutilized because of a lack of relevant information on available technologies. One way to provide information to a large population is through social networks. Governments and communities often disseminate energy-saving information to occupants, for example, as it can be spread very efficiently through local social networks (Stern 1992). Recently, a series of empirical studies have suggested that exposing occupants to energy consumption information from members in their peer networks can be an effective way of changing energy conservation decisions and

behaviors (Peschiera, Taylor, and Siegel 2010; Peschiera and Taylor 2012). These studies report that users who were exposed to electricity use information from their peer network achieved a statistically significant drop in electricity consumption as compared to users who did not have access to energy usage information of their peer network. Their data also strongly suggested that the structural properties of peer networks were positively correlated with energy conservation. Additionally, simulation studies have also been conducted to understand changes to building occupants' energy consumption behavior under different peer network structures (Anderson, Lee, and Menassa 2012; Azar and Menassa 2012; Chen, Taylor, and Wei 2012). Some components of network structure analyzed in these studies include network type, size, degree of each node, and social influence from one's peers. While these studies are of offline peer networks, similar phenomena may occur in online social networks.

Network structures of social networking sites like Facebook, Twitter and Google+, are distinguished from traditional ones by their sizes, connectivity, and dynamics. Online social networks connect millions of users and aggregate into massive structures. The ease of connecting with another user allows formation of large personal networks which is practically impossible in the offline world. Also, as new users continue to join daily and form new connections, the networks are continuously growing and evolving. Compared to traditional offline networks, online social networks allow information to reach a large population at a much lower cost and with far fewer barriers (Kwak et al. 2010).

Therefore, it is important to gain a proper understanding of how information, specifically information related to energy conservation, travels through large online social networks. This will aid governments and organizations to most effectively leverage online networks to change user behavior and reduce overall energy consumption.

2   BACKGROUND

Researchers have studied word of mouth (WOM) communication for decades. WOM plays an important role in the diffusion of innovation and information (Mahajan, Muller, and Bass 1990). Traditionally, WOM was an oral, person-to-person communication method (Arndt 1967). A communicator and a recipient exchanged some ephemeral oral or spoken messages in WOM communication (Stern 1994). The term *information cascade* refers to the causal propagation of information (Leskovec, Adamic, and Huberman 2007) emerging from micro-level WOM communication. With the development of information technologies in the last two decades, information cascades can also occur in other communication channels, such as mobile device messaging, online chatting and through social networks (Chan and Ngai 2011). In online social networking platforms, information propagates through a social network via WOM information exchanges (Cha, Mislove, and Gummadi 2009). A communicator passes a piece of information to recipients in his/her social network. The recipients then become communicators and pass the information to their connections.

Research has shown that many factors affect the level of information propagation, i.e. how many levels beyond the information originator the information can reach, including: (1) communicator's influence (Chan and Ngai 2011); (2) homophily (Aral, Muchnik, and Sundararajan 2009); (3) tie strength (Brown and Reingen 1987; Goldenberg, Libai, and Muller 2001); (4) recipient's perception (Bakshy et al. 2011; Chan and Ngai 2011); (5) network density (Webster and Morrison 2004; Vilpponen, Winter, and Sundqvist 2006); and (6) the number of an initiator's connections (Wu and Huberman 2004; Bakshy et al. 2011). There is a large amount of literature exploring the impact of the first four factors, the last two—network density and the number of an initiator's connections—have not attracted as much scholarly attention. Network density and the number of an initiator's connections may operate differently in large-scale online networks, because of the new structures of online social networks. These two factors are described in the next two subsections.

2.1 Network Density

The first factor is *network density*, which represents the ratio of links that are present in a network over the total number of possible links (Wasserman and Faust 1994). Density can be calculated using the following formula for an undirected network:

$$\Delta = \frac{2L}{g(g-1)} \tag{1}$$

Where Δ is the density, L is the number of present links, and g is the number of nodes in the network.

Online social networks often have lower density than offline networks. Although online networks have large populations of users, they tend to form relatively few connections. In a study of marketing networks, Webster and Morrison (2004) found the density of the network was only 3%. Vilpponen, Winter, and Sundqvist (2006) analyzed blog re-postings through electronic word-of-mouth communication and found the density to be only 0.6% in a network of 360 blogs. It is intuitive to assume that the higher the density, the further information can travel because each node has the potential to pass information to more recipients if he/she has a large number of connections. However, there is a lack of evidence as to whether network density can significantly influence information cascades. Additionally, the relative importance of network density as compared to other factors is unknown.

2.2 Number of an Initiator's Connections

In online social networking platforms, researchers often use the number of connections of a node to measure its influence, with the assumption that an influential member will propagate information better (Cha, Mislove, and Gummadi 2009; Kwak et al. 2010). However, there is some controversy over whether the number of an initiator's connections does in fact have an impact on information cascades. Some studies have found that well-connected nodes positively support information propagation. If highly connected individuals are provided with a particular opinion, they can be very effective in distributing this opinion throughout their network over the long term (Wu and Huberman 2004). Bakshy and colleagues (2011) also found that the largest cascades of information tended to be generated by Twitter users with a large number of followers.

These conclusions have been challenged by the findings of other studies. Watts and Dodds (2007) conducted a simulation-based study and concluded that important and influential nodes had only a limited impact on the diffusion of innovation. Leskovec and colleagues (2007) reported a similar effect and found that the highly connected nodes had a limited influence in a recommendation network because the success of their recommendations declined quickly as they continued to make recommendations.

These contradictory findings raise the question of whether intervention strategies for behavioral change should focus on special individuals with a high number of connections or not. Combined with the limited evidence about the effects of network density on the level of information cascades, there is a need for better understanding of the new networks structures arising from large online social networks, especially of their impact on disseminating information. In the work reported here, we simulated how online social networks have been used to spread energy saving information.

3 METHODOLOGY

3.1 Development of Hypotheses

Existing studies have demonstrated that density can influence information cascades in online social networks (Webster and Morrison 2004; Vilpponen, Winter, and Sundqvist 2006). Therefore, we proposed the following hypothesis:

    Hypothesis 1a: Network density strongly influences information cascades.

Also, very little research has examined the relative importance of network density compared to other factors, for example homophily or communicator's influence. Chen, Taylor, and Wei (2012) studied offline occupant networks, and found that network degree, which is tightly related to network density, influences an information recipient's behavior much more strongly than other factors. Anderson, Lee, and Menassa (2012) also found that increased social connectivity could significantly decrease the time for occupants to adopt energy-conservative behaviors. These studies provide evidence that network density could be more influential than other factors in increasing the scale of information cascades. Thus, another hypothesis was proposed as follows:

Hypothesis 1b: Network density is more influential than other factors.

Research has found that influential nodes facilitate deep information cascades in social networks (Lerman and Ghosh 2010), although this has been challenged by others (Cha, Mislove, and Gummadi 2009; Kimura et al. 2010; Kwak et al. 2010). In online social networks, an influential node generally has a large number of connections, so he/she has a higher probability of passing information on to a larger population, and thus generates a large propagation. Thus, we tested the following hypothesis:

Hypothesis 2a: The number of an initiator's connection strongly influences information cascades.

Very little research has studied the relative importance of the number of an initiator's connection to other factors that influence information cascades. Existing studies have found that besides an influential node, a critical mass of easily influenced audiences may be just as important (Watts and Dodds 2007; Bakshy et al. 2011). However, no quantitative values have been reported that would allow us to rank them. Studies of cascade models have found that a small set of influential nodes can generate large cascades (Goldenberg, Libai, and Muller 2001; Kempe, Kleinberg, and Tardos 2003). Therefore, based on our belief that the number of an initiator's connections is more influential than other factors, we proposed the following hypothesis:

Hypothesis 2b: The number of an initiator's connection is more influential than other factors.

3.2 Development of Information Cascade Model

Based on the literature discussed above, an information cascade model was developed using an agent-based modeling approach (Figure 1). The Python programming language was used to implement the algorithms.

First, a scale-free network based on Barabasi and Albert's function was generated (equation 2) (Barabasi 2002).

$$G = \text{barabasi\_albert\_graph}(N, m) \qquad (2)$$

Where G is the generated graph, N the number of nodes and m the number of edges attaching each new node to the existing nodes. This function was implemented using *networkx*, a Python package. Each node in the network represents a user and each user has 4 attributes, defined in Table 1: homophily, tie strength, communicator's influence, and recipient's perception. A node N_j was randomly selected and its immediate connections identified, denoted as C_{Nj}. For each connection $C_{Nj}(i)$, if the difference in homophily between N_j and $C_{Nj}(i)$ was greater than 0.1 (Aral, Muchnik, and Sundararajan 2009), information could not pass, and the model went back to test another connection $C_{Nj}(i+1)$. If the difference of homophily was less than 0.1, the social influence *SI* was calculated using N_j's communicator's influence, $C_{Nj}(i)$'s recipient's perception, and the tie strength between N_j and $C_{Nj}(i)$. If *SI* was larger than 1.825 (Goldenberg, Libai and Muller 2001), information passed from N_j and $C_{Nj}(i)$ and $C_{Nj}(i)$ became the next N_j. This process was repeated 50 times.
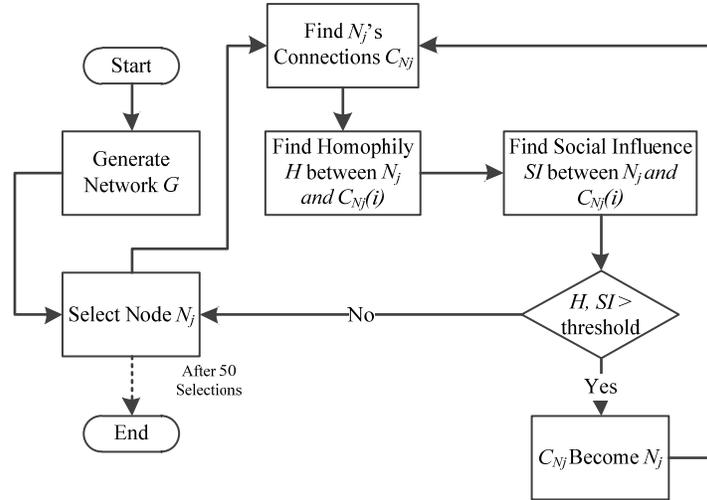
Figure 1: Algorithm Schematic for the Information Cascade Model

To maximize computational efficiency, we set N to 1,500 in the base model and simulated a relatively large network. When information cascades in this network, it will travel approximately 15 steps from the information originator to the end recipient, which significantly exceeds the value of steps we need to test our hypotheses. We set m to 1, the lower boundary that equation 2 can take. This gives a network density for the generated G of 0.00133. Thirty different random seeds were used to generate 30 different structures with the same N and m, leading to a model simulation total of 1500 (50×30) times. The propagation depth, defined as the steps information traveled, was recorded as the output.

Table 1: Definition of Input Parameters and Their Distributions for Variance-Based GSA

Input Parameter	Definition	Distribution
Homophily	The degree of similarity between a dyad of nodes (Aral, Muchnik, and Sundararajan 2009)	$U^* = (1, 3)$
Tie Strength	The intensity of the social relation between a dyad of nodes (Brown and Reingen 1987)	$U = (0, 0.5)$
Communicator's Influence	The influence of the communicator providing the information (Chan and Ngai 2011)	$U = (0, 0.5)$
Recipient's Perception	The ease with which a recipient is influenced (Bakshy et al. 2011)	$U = (0, 0.5)$
Network Density	Ratio of links that are present in a network over the total possible links (Wasserman and Faust 1994)	$D^{**} = (0.001, 0.65)$
The Number of an Initiator's Connections	The number of connections an initiator possesses (Leskovec, Adamic, and Huberman 2007; Cha, Mislove, and Gummadi 2009).	$D = (1, 450)$

*Uniform distribution; **Discrete distribution

3.3   Variance-Based Global Sensitivity Analysis (GSA)

A computational experiment was designed to determine how influential each factor was and reveal the relative importance of one factor to another. The experiment utilized the variance-based GSA method, which made it possible to decompose the output, and thus show how the output variance depends on the input factors (Saltelli et al. 2008; Ligmann-Zielinska and Sun 2010), enabling us to rank the parameters in

order of their importance (Saltelli et al. 2008). This same approach was used previously by other researchers conducting uncertainty analysis on agent-based models (Ligmann-Zielinska and Sun 2010).

For the variance-based GSA, two indicators were calculated: the first order sensitivity index S_i and the total order sensitivity index S_{Ti}. A high S_i indicates that parameter i values could themselves substantially influence output uncertainty, so i is clearly an important factor for information cascades. A high S_{Ti} means that the combined influence of the parameter i and its interactions with all the other parameters is substantial (Saltelli et al. 2008). If a parameter has low values for both S_i and S_{Ti}, this would be deemed an unimportant factor in determining the information propagation depth.

Table 1 shows how distributions were assigned for the model input. Monte Carlo simulation was employed and the model executed 40,000 times. The information propagation depth was recorded as the output.

4   RESULTS

4.1   Results from Base Model Simulation

The simulation results from the base model showed that when network density was around 0.00133, the depth of information was around 2.26.

4.2   Results from Variance-Based Global Sensitivity Analysis

Table 2 shows the values of the first order sensitivity index S_i and the total order sensitivity index S_{Ti}. S_i for network density was less than 0.015, which indicates that the factor itself contributed less than 1.5% to the simulation output. Compared to the other factors, it contributed the least to the variance. S_i for the number of an initiator's connections was about 0.08, which means that the number of an initiator's connections contributed 8% to the output. While the value was lower than the S_i's of other factors, although it was higher than that for the network density.

Table 2: First Order Sensitivity Index S_i and Total Order Sensitivity Index S_{Ti} of Parameters

Parameters	Homophily	Tie Strength	Communicator's Influence	Recipient's Perception	Network Density	No. of an initiator's connections
S_i	0.288	0.109	0.103	0.114	0.0146	0.0807
S_{Ti} (Normalized)	0.203	0.0566	0.00358	0.0472	0.413	0.276

5   DISCUSSION

These results confirmed that all 6 factors played some role in influencing information cascades in online social networks. However, there were some interesting differences in the patterns between the factors that have increased our understanding of how information cascades more generally through online social networks. The following section discusses each of our hypotheses in the order presented in Section 3.1 above.

In our first hypothesis, based on existing studies (Webster and Morrison 2004; Vilpponen, Winter, and Sundqvist 2006), we assumed network density would be a strongly influential factor. However, the value of S_i was less than 0.015, which indicates that network density alone contributes only 1.5% to information propagation increases. Therefore, Hypothesis 1a is rejected.

For Hypothesis 1b, based on reports in the literature of research into offline networks (Chen, Taylor, and Wei 2012; Anderson, Lee, and Menassa 2012), we expected network density to have more influence than the other 4 factors. However, the value of S_i was much lower than any other input factors and this hypothesis was therefore also rejected. Even if other online social networks had higher densities, the model predicts that energy saving information would fail to propagate deeply.

Again based on the literature (Lerman and Ghosh 2010), Hypothesis 2a proposed that the number of an initiator's connections would strongly influence information cascades. However, as the data in Table 2 shows, the value of S_i was only about 0.08, which means that the number of an initiator's connections only influenced the model output by 8%. Therefore, Hypothesis 2a was also rejected. This suggests that even if opinion leaders or celebrities were recruited to pass on energy saving information via their social networking platforms, the penetration would still be limited. Although the information would reach their immediate connections, i.e. their followers, there would be only a low chance that the information would penetrate more deeply.

In Hypothesis 2b, we proposed that the number of an initiator's connection would influence information cascades more than the other 4 factors based on previous reports in the literature (Goldenberg, Libai, and Muller 2001; Kempe, Kleinberg, and Tardos 2003). However, the value of S_i was about 0.08, and although this was higher than the S_i for network density, it was still lower than other factors. Therefore, Hypothesis 2b was rejected.

Our findings demonstrate that, for online social networks, cascades of energy saving information: (1) failed to propagate deeply; and (2) are not strongly impacted by either network density or the number of an initiator's connections.

6   CONCLUSIONS

Online social networking platforms can potentially pass energy saving information to a large population, but this study is the first to focus specifically on this subject. Existing research has reported contradictory findings regarding whether network density and the number of an initiator's connection influence information cascades, so a simulation study was conducted to quantitatively calculate the impacts of a range of network attributes using an agent-based information cascade model. A variance-based GSA experiment was also performed.

We found that the two factors have only a limited impact on information cascades: the combined influence of network density and the number of an initiator's connection is less than 10%. Therefore, although the new media provides an innovative way to contact a large population at very little cost and with a low barrier, it still faces the challenge of diffusing energy saving information to a mass audience. Even the existence of massive networks and a large number of potential recipients cannot guarantee deep cascades of energy saving information. To fully utilize the potential of online social networking platforms, the information needs to be intriguing, the communicators need to be knowledgeable, credible, and passionate, and the audiences need to have sufficient background knowledge to be interested and to want to get involved.

REFERENCES

Anderson, K., Lee, S. and Menassa, C. (2012). Effect of social network type on building occupant energy use. *Proceedings of the Fourth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*. Toronto, Ontario, Canada, ACM: 17-24.

Aral, S., Muchnik, L. and Sundararajan, A. (2009). "Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks." *Proceedings of the National Academy of Sciences,* 106(51): 21544-21549.

Arndt, J. (1967). "Role of product-related conversations in the diffusion of a new product." *Journal of Marketing Research,* 4(3): 291-295.

Azar, E. and Menassa, C. (2012). "Agent-based modeling of occupants and their impact on energy use in commercial buildings." *Journal of Computing in Civil Engineering,* 26(4): 506-518.

Bakshy, E., Hofman, J. M., Mason, W. A. and Watts, D. J. (2011). Everyone's an influencer: Quantifying influence on twitter. *Proceedings of the fourth ACM international conference on Web search and data mining*. Hong Kong, China, ACM**:** 65-74.

Barabasi, A. L. (2002). "Linked: The new science of networks. 2002." *Cambridge, MA: Perseus*.

Brown, J. J. and Reingen, P. H. (1987). "Social ties and word-of-mouth referral behavior." *Journal of Consumer Research,* 14(3): 350-362.

Cha, M., Mislove, A. and Gummadi, K. P. (2009). A measurement-driven analysis of information propagation in the flickr social network. *Proceedings of the 18th international conference on World wide web*. Madrid, Spain, ACM**:** 721-730.

Chan, Y. Y. Y. and Ngai, E. (2011). "Conceptualising electronic word of mouth activity: An input-process-output perspective." *Marketing Intelligence & Planning,* 29(5): 488-516.

Chen, J., Taylor, J. E. and Wei, H.-H. (2012). "Modeling building occupant network energy consumption decision-making: The interplay between network structure and conservation." *Energy and Buildings,* 47(0): 515-524.

Energy Information Administration. (2011). "Annual energy review." *Energy Information Administration, US Department of Energy: Washington, DC www. eia. doe. gov/emeu/aer*.

Goldenberg, J., Libai, B. and Muller, E. (2001). "Talk of the network: A complex systems look at the underlying process of word-of-mouth." *Marketing letters,* 12(3): 211-223.

Kempe, D., Kleinberg, J. and Tardos, É. (2003). Maximizing the spread of influence through a social network. *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM.

Kimura, M., Saito, K., Nakano, R. and Motoda, H. (2010). "Extracting influential nodes on a social network for information diffusion." *Data Min Knowl Disc,* 20(1): 70-97.

Kwak, H., Lee, C., Park, H. and Moon, S. (2010). What is twitter, a social network or a news media? *Proceedings of the 19th international conference on World wide web*, ACM.

Lerman, K. and Ghosh, R. (2010). Information contagion: An empirical study of the spread of news on digg and twitter social networks. *Proceedings of 4th International Conference on Weblogs and Social Media (ICWSM)*.

Leskovec, J., Adamic, L. A. and Huberman, B. A. (2007). "The dynamics of viral marketing." *ACM Trans. Web,* 1(1): 5.

Ligmann-Zielinska, A. and Sun, L. (2010). "Applying time-dependent variance-based global sensitivity analysis to represent the dynamics of an agent-based model of land use change." *International Journal of Geographical Information Science,* 24(12): 1829-1850.

Mahajan, V., Muller, E. and Bass, F. M. (1990). "New product diffusion models in marketing: A review and directions for research." *Journal of Marketing,* 54(1): 1-26.

Masoso, O. T. and Grobler, L. J. (2010). "The dark side of occupants' behaviour on building energy use." *Energy and Buildings,* 42(2): 173-177.

National Science Foundation. (2009). "Building a sustainable energy future u.S. Actions for an effective energy economy transformation." from http://purl.access.gpo.gov/GPO/LPS118254.

Peschiera, G., Taylor, J. E. and Siegel, J. A. (2010). "Response–relapse patterns of building occupant electricity consumption following exposure to personal, contextualized and occupant peer network utilization data." *Energy and Buildings,* 42(8): 1329-1336.

Peschiera, G. and Taylor, J. E. (2012). "The impact of peer network position on electricity consumption in building occupant networks utilizing energy feedback systems." *Energy and Buildings,* 49(0): 584-590.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. and Tarantola, S. (2008). *Global sensitivity analysis: The primer*, Wiley-Interscience.

Stern, B. B. (1994). "A revised communication model for advertising: Multiple dimensions of the source, the message, and the recipient." *Journal of Advertising,* 23(2): 5-15.

Stern, P. C. (1992). "What psychology knows about energy conservation." *American Psychologist; American Psychologist,* 47(10): 1224.

Turner, C. and Frankel, M. (2008). *Energy performance of leed for new construction buildings*, New Buildings Institute Washington, DC.

Vilpponen, A., Winter, S. and Sundqvist, S. (2006). "Electronic word-of-mouth in online environments: Exploring referral network structure and adoption behavior." *Journal of Interactive Advertising,* 6(2): 71-86.

Wasserman, S. and Faust, K. (1994). *Social network analysis : Methods and applications*. Cambridge; New York, Cambridge University Press.

Watts, D. J. and Dodds, P. S. (2007). "Influentials, networks, and public opinion formation." *Journal of Consumer Research,* 34(4): 441-458.

Webster, C. M. and Morrison, P. D. (2004). "Network analysis in marketing." *Australasian Marketing Journal (AMJ),* 12(2): 8-18.

Wilson, C. and Dowlatabadi, H. (2007). "Models of decision making and residential energy use." *Annu. Rev. Environ. Resour.,* 32: 169-203.

Wu, F. and Huberman, B. A. (2004). Social structure and opinion formation. *Computational Economics 0407002, EconWPA*.

AUTHOR BIOGRAPHIES

QI WANG is a PhD student and Pratt Fellowship recipient in the Charles E. Via, Jr. Department of Civil and Environmental Engineering at Virginia Tech. He is conducting simulation research on the NSF-funded "Building Occupant Network Dynamics" CAREER project. His simulation work focuses on understanding and modeling energy saving information cascades in online social networks. His email address is wangqimsu@gmail.com.

JOHN E. TAYLOR is an Associate Professor in the Charles E. Via, Jr. Department of Civil and Environmental Engineering at Virginia Tech. He received his Ph.D. from Stanford University in 2006. He is Director of the Civil Engineering Network Dynamics Lab at Virginia Tech and specializes in the investigation of civil engineering network dynamics of industrial and societal importance. His current research focuses on three key network dynamics; (1) achieving sustained energy conservation in the built environment by coupling energy use and energy conservation practices with building occupant networks and across buildings in building networks, (2) the trending increase in global outsourcing of complex engineering services, and (3) examining the impact of integrated information systems on project networks and the associated virtualization of the engineering workforce. His email is <jet@vt.edu>